SUBSTITUTE SPEC - REDLINED VERSION

ESSENTIAL GENES FROM C. ALBICANS AND A METHOD FOR SCREENING ANTIMYCOTIC SUBSTANCES USING SAID GENES

Field of the Invention

The present invention relates to a method for screening for antimycotic substances in which essential genes from mycetes, particularly from Candida albicans (C. albicans) as well as functionally similar genes from other pathogenic mycetes, or the corresponding encoded proteins, are used as targets. The invention also relates to specific C. albicans genes.

Background of the Invention

The spectrum of known fungal infections stretches from fungal attack of skin or nails to potentially hazardous mycotic infections of the inner organs; Such infections and resulting diseases are known as mycosis.

Antimycotic substances (fungistatic or fungicidal) are used for treatment of mycosis. However, up to now, relatively few substances with pharmacological effects are known, such as Amphotericin B, Nystatin, Pimaricin, Griseofulvin, Clotrimazole, 5-fluoro-cytosine and Batraphene. The drug treatment of fungal infections is extremely difficult, in particular because both the host cells and the mycetes, are eucaryotic cells. Administration of drugs based on known antimycotic substances results therefore often in undesired side-effects, for example Amphotericin B has a nephrotoxic effect. Therefore, there is a strong need for pharmacologically efficient substances usable for the preparation of drugs, which are suitable for prophylactic treatments of immunodepressive states or for the treatment of an existing fungal infection. Furthermore, the substances should exhibit a specific spectrum of action in order to selectively inhibit the growth and proliferation of mycetes without affecting the treated host organism.

Summary of the Invention

The aim of the present invention is to provide a method for the identification of antimycotic substances and especially for the identification of anti-Candida substances. An essential feature of this method is that essential genes from mycetes are used as targets for the screening.

The present invention thus concerns a method for screening antimycotic substances wherein an essential gene from mycetes or a functionally similar gene in another pathogenic mycete, or the corresponding encoded protein, is used as target and wherein the essential gene is selected from the group consisting of CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 and CaJL039.

Brief Description of the Figures

Figure 1 shows the comparison of the deduced protein of YNL256 gene with YNL256.

Figure 2 shows the comparison of the deduced YBR102 gene with YNL102.

Figure 3 shows the alignment of ScOR110-homologous fragment with Candida albicans CaOR110 sequence.

Figure 4 is a map of CaOR110.

Figure 5 is a map of the CaOR110 variant.

Figures 6A, 6B and 6C show the alignment of the original CaOR110 and splice variant.

Figure 7 shows a map of CaDR325.

Detailed Description of the Invention

According to one embodiment of the method of the invention mycete cells which express the essential gene, or a functionally similar mycete gene, to a different level are incubated with the substance to be tested and the growth inhibiting effect of the substance is determined.

According to another embodiment, said target gene or the corresponding target gene encoded protein is contacted in vitro with the substance to be tested and the effect of the substance on the target is determined.

According to another embodiment, the screened substances inhibit partially or totally the functional expression of the essential genes or the functional activity of the encoded proteins.

According to one embodiment the screened substances partially or totally inhibit the activity of dihydropteroate synthase (DHPS) and/or 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK).

According to another embodiment, the mycete species are selected from the group comprising Basidiomycetes, Ascomycetes and Hyphomycetes.

According to another embodiment of the method of the invention said functional similar genes are essential genes from Candida Spp., preferably Candida albicans, or from Aspergillus Spp., preferably from Aspergillus fumigatus.

According to another aspect, the present invention concerns a polynucleotide having the sequence as depicted in SEQ ID No. 2, SEQ ID No. 5~~4~~, SEQ ID No. 8~~6~~, SEQ ID No. 10~~7~~, SEQ ID No. 13~~9~~, SEQ ID No. 15~~10~~, SEQ ID No. 17~~11~~ or SEQ ID No. ~~13~~20, preferably SEQ ID No. 2, SEQ ID No. 5~~4~~, SEQ ID No. 8~~6~~, SEQ ID No. 13~~9~~, SEQ ID No. 14~~10~~ or SEQ ID No. 17~~11~~, homologs thereof and functional fragments thereof.

According to another aspect, the present invention concerns a gene which is CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, preferably CaOR110, CaMR212, CaNL256, CaBR102 or CaIR012, or a functionally similar gene or a functional fragment thereof.

According to this embodiment, the functionally similar gene or homologous polynucleotide has a sequence identity, at the nucleotide level, with CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, respectively, of at

least 50%, preferably of at least 60%, and most preferably of at least 70%. A functional fragment is a polynucleotide fragment that will retain the functionality of the starting product (nucleotide or gene). One example is the CaOR110 splice variant (which is also homologous to the original gene, with about 90% identity).

According to another embodiment, the functionally similar gene has a sequence identity, at the amino-acid level, with CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, respectively, encoded protein (s) of at least 40%, preferably of at least 50%, more preferably of at least 60% and most preferably of at least 70%.

These figures given for the gene apply mutatis mutandis to the polynucleotide, as far as homology and similarity.

According to another aspect, the present invention covers the protein(s) encoded by CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, respectively, gene(s) or by a functionally similar gene, or a functional polypeptidic fragment thereof.

According to another aspect, the present invention provides a plasmid containing CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, respectively, gene (s), a functionally similar gene or a functional fragment thereof According to another aspect, the present invention provides a plasmid (bacteria containing same) deposited at the CNCM (Institut Pasteur, Paris) on 98/08/13, with the accession numbers 1-2065, I-2063 and I-2064, corresponding to the CaNL256, CaBR102 and CaIR012 genes, respectively.

According to another aspect, the present invention provides a plasmid (bacteria containing same) deposited at the DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany) on 99/08/06 with the accession numbers DSM 12977, DSM 12976, DSM 12978 and DSM 12979, corresponding to the CaDR325, CaOR110, CaOR110 splice variant and CaMR212, respectively.

According to another aspect, the present invention provides a kit for diagnosis of fungal infections comprising a gene selected from the group consisting of CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 and CaJL039, a functionally similar gene thereof, a functional fragment thereof, the corresponding encoded protein or a functional polypeptide fragment thereof.

According to another aspect, the present invention provides an antibody directed against the protein encoded by the CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, respectively, gene(s) or by a functionally similar gene, or a polypeptide fragment thereof.

According to another aspect, the present invention provides a polynucleotide obtainable by the process comprising the following steps:

(i) selecting an essential gene from Saccharomyces cerevisiae;

(ii) comparing the sequence of said gene with Candida Albicans genome sequences;

(iii) deducing ~~homologuous~~ homologous oligonucleotides regions ;

(iv) PCR amplifying the thus-obtained oligonucleotides;

(v) using the amplimers of step (iv) for detecting the complete gene of interest:

the amplimers of step (iv) are used as a probe for detecting the complete gene of interest from a Candida albicans genomic or cDNA library; or

the complete gene is obtained by 3'and 5'extension of the amplimer, e. g. by using a PCR method.

According to the invention, the first step is to identify said essential genes and starting from these thus identified genes, essential genes from other pathogenic mycetes can be identified. For practical purposes, essential genes from S. cerevisiae are first identified and

starting from them, essential genes from other pathogenic fungus, especially from Candida, are obtained.

The present invention thus discloses the identification of essential genes from C. albicans and their use in a method for the screening of antimycotic substances, especially anti-Candida substances.

In order to identify essential genes of S. cerevisiae, individual genomic genes are eliminated through homologous recombination. If the DNA segment thus eliminated concerns an essential gene, then the deletion is lethal for the S. cerevisiae cells in haploid form.

A method, wherein the studied S. cerevisiae gene is replaced by a marker gene can be used to generate the corresponding genomic deletion of S. cerevisiae and to determine the S. cerevisiae cells containing the deletion.

As a selection marker a dominant selection marker (e. g. kanamycin resistance gene) or an auxotrophic marker can for example be used. As an auxotrophic marker, it is possible to use genes coding for key enzymes of amino acid or nucleic base synthesis. For example, one can use as a selection marker the following genes from S. cerevisiae : gene encoding for the metabolic pathway of leucine (e. g. LEU2-gene), histidine (e. g. HIS3-gene) or tryptophan (e. g. TRP-1 gene) or for the nucleic base metabolism of uracil (e. g. URA3-gene).

Auxotrophic S. cerevisiae strains can be used. These auxotrophic strains can only grow on nutritive media containing the corresponding amino acids or nucleotide bases. All laboratory S. cerevisiae strains, containing auxotrophic markers can for instance be used. When diploid S. cerevisiae strains are used, then the corresponding marker gene must be homozygously mutated. Strain CEN. PK2 or isogenic derivatives thereof can be used.

Strains containing no suitable auxotrophic marker can also be used such as prototrophic S. cerevisiae strains.Then a dominant selection marker e. g. resistance

gene, such as kanamycin resistance gene can be used. A loxP-KanMX-loxP cassette can advantageously be used for this purpose.

For the homologous recombination replacing the whole DNA sequence or part thereof of a S. cerevisiae gene, DNA fragments are used wherein the marker gene is flanked at the 5'-and 3'-ends by sequences which are homologous to the 5'-and 3'-ends of the studied S. cerevisiae gene.

Different processes can be used for the preparation of the corresponding DNA fragments which are also appropriate for the deletion of any specific S. cerevisiae gene. A linear DNA-fragment is used for the transformation of the suitable S. cerevisiae strain. This fragment is integrated into the S. cerevisiae genome by homologous recombination. These processes include:

1."Conventional method"for the preparation of deletion cassettes (Rothstein, R. J. (1983) Methods in Enzymology, Vol. 101,202-211).

2."Conventional Method"using the PCR technique ("modified conventional method").

3. SFH (short flanking homology)-PCR method (Wach, A. et al. (1994) Yeast 10: 1793-1808; Gültner, U. et al. (1996) Nucleic Acids Research 24: 2519-2524).

1. In the"conventional method"for the preparation of deletion cassettes in the S. cerevisiae genome, the gene to be studied is either already present in an appropriate vector or is integrated in such a vector. With this method, any pBR-pUC-and [pBluescript®-derivatives] can be used for example. A major part of the target gene sequence is eliminated from the vector, for instance using appropriate restriction sites, conserving however the 3'-and 5'-regions of the studied gene inside the vector. The selected marker gene is integrated between the remaining regions.

2. In the modified form of this "conventional method", PCR is used. This method allows amplification of the 3'- and 5'-terminal regions of the coding sequence of the

studied S. cerevisiae gene. This method amplifies selectively both terminal regions of the studied gene, therefore, two PCR-reactions must be carried out for each studied gene, amplifying once the 5'-end, and once the 3'-end of the gene. The length of the amplified terminal DNA-fragment depends on the existing restriction sites. The amplified terminal ends of the studied gene have generally a length of 50 to 5000 base pairs (bp), preferably a length between 500 and 1000 bp.

As template for the PCR-reactions, genomic DNA of S. cerevisiae or wild-type genes can be used. The primer-pairs (a sense and an antisense primer, respectively) are constructed so that they correspond to the 3'-end and the 5'-end sequence of the studied S. cerevisiae gene. Especially, the primer is selected such as to allow its integration by way of appropriate restriction sites.

As vectors, pBR-pUC-and pBluescript®-derivatives can be used. In particular vectors already containing a gene encoding the selection marker, are appropriate. In particular, vectors can be used, which contain genes of the selection marker HIS3, LEU2, TRP1 or URA3.

The DNA segments of the studied S. cerevisiae gene, obtained by PCR, are integrated in the vector at both sides of the selection marker, so that subsequently, as in the "conventional method", the selection marker is flanked on both ends by DNA sequences which are homologous to the studied gene.

3. Homologous recombination in S. cerevisiae takes place in a very efficient and precise manner and the length of the DNA sequence homologous to the studied S. cerevisiae gene flanking the selection marker gene can in fact be considerably shorter than with the "modified conventional method". The flanking ends homologous to the studied S. cerevisiae gene need to present a length of only about 20-60 bp, preferably 30-45 bp. The SFH-PCR method is

particularly advantageous as the laborious cloning step can be obviated.

A PCR reaction is carried out on a DNA-template containing the gene for the selection marker to be used, wherein the primers are constructed so that the DNA sequence of the sense primer is homologous to the 5'-end of the selection marker sequence and so that the primer presents in addition at its 5'-end a region of preferably 40 nucleotides, which corresponds to the 5'-terminal sequence of the studied S. cerevisiae gene. The antisense primer is constructed in an analogous manner, i. e. it is complementary to the 3'-end of the gene sequence of the selection marker, wherein this primer contains at its 5'-end a region of also preferably 40 nucleotides, which corresponds to the complementary strand of the 3'-terminal coding sequence of the studied gene.

For the amplification of S. cerevisiae genes to be studied by the SFH-PCR method, vectors containing the gene for the auxotrophic marker or selection marker can be used. Especially, plasmid pUG6 is used as the template. This plasmid contains a loxP-KanMX-loxP cassette (Gültner, U. et al. (1996) Nucleic Acids Research 24: 2519-2524). In other terms, the Kanamycin resistance gene is flanked at both ends by a loxP sequence (loxP-KanMX-loxP cassette). This cassette is advantageous in that the Kanamycin resistance gene can be eventually eliminated from the S. cerevisiae genome after integration of the loxP-KanMX-loxP cassette into the S. cerevisiae gene to be studied. Cre-recombinase of bacteriophage P1 can be used for this purpose. Cre-recombinase recognizes the loxP sequences and induces elimination of the DNA located between the two loxP sequences by a homologous recombination process. As a result only one loxP sequence remains and the so-called marker regeneration occurs, i. e. the S. cerevisiae strain may be transformed again using the loxP-KanMX-loxP cassette. This is particularly advantageous, when at least

two functionally similar genes are to be deleted in order to obtain a lethal phenotype.

With the PCR-method, the PCR reaction primers are at the 3'-end a preferably 20 nucleotide long sequence, which is homologous to the sequence situated left and/or right of the loxP-KanMX-loxP cassette, and at the 5'-end a preferably 40 nucleotide long sequence, which is homologous to the terminal ends of the gene to be studied.

Using the three methods, one obtains linear deletion cassettes containing the gene encoding the selection marker, which is flanked on both sides by homologous sequences of the gene to be studied. The deletion cassettes are used for the transformation of diploid S. cerevisiae strains. The diploid strain S. cerevisiae CEN. PK2 (Scientific Research & Development GmbH, Oberursel) can be used for example for this purpose.

[CEN. PK2 Mata/MAT a ura3-52/ura3-52 leu2-3,112/leu2-3,112his3Al/his3Al trpl-289/trpl-289 MAL2-8C/MAL2-8c SUC2/SUC2]

The strain CEN. PK2 is prepared and cultivated using known methods (Gietz, R. D. et al. (1992) Nucleic Acids Research 8: 1425; Guldener, U. et al. (1996) Nucleic Acids Research 24: 2519-2524).

The cells of the S. cerevisiae strain used are transformed according to known processes with an appropriate DNA quantity of the linear deletion cassette (e. g. Sambrook et al. 1989). Thereafter, the medium in which the cells are cultivated is replaced by a new medium, a so-called selective medium, which does not contain the corresponding amino acid (e.g. histidine, leucine or tryptophan) or nucleic base (e.g. uracil) or, when using a deletion cassette containing the kanamycin resistance gene, by a medium containing geneticin (G4180) (e. g. a complete medium (YEPD) containing geneticin). Alternatively, the transformed cells may be plated on agar plates prepared using the corresponding media. Thereby, one selects the

transformed cells, in which a homologous recombination occured, since only those cells can grow under these modified conditions.

However, in most cases, only one of the two copies of the gene in the double chromosome set of a diploid S. cerevisiae strain is replaced by the DNA of the deletion cassette during the transformation, resulting in a heterozygote-diploid S. cerevisiae mutant strain, wherein one copy of the gene studied is replaced by a selection marker, while the other copy of the gene is maintained in the genome. This presents the advantage that in case of a deletion of an essential gene, due to the existence of the second copy of the essential gene, the mutant S. cerevisiae strain is still viable.

The proper integration of the deletion cassette DNA at the predetermined chromosomal gene locus (gene locus of the gene to be studied) may be checked by Southern-Blot Analysis (Southern, E. M. (1975) J. Mol. Biol. 98: 503-517) or by diagnostic PCR analysis using specific primers (Guldener, U. et al. (1996) Nucleic Acids Research 24: 2519- 2524)

The genetic separation of individual diploid cells may be monitored by tetrad analysis. To this end, reduction division (meiosis) is induced in the diploid cells, especially heterozygote mutant strains, using known methods such as nitrogen impoverishment on potassium acetate plates (Sherman, F. et al. (1986) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.; Guthrie, C. and Fink, G. R. (1991) Methods in Enzymology, Vol 194. Academic Press, San Diego, 3-21; Ausubel, F. M. et al. (1987) Current Protocol in Molecular Biology John Wiley and Sons, Inc., Chapter 13). Meiosis results only in asci with four ascospores (segregated), which can be indivualized after partial enzymatic digestion of the ascospore wall with zymolyase (Ausubel et al. (1987)) by way of micromanipulators (e. g. SINGER). For example when a tetrad

analysis is carried out on a heterozygote-diploid mutant strain in which an essential gene present in the double chromosome set is replaced on one chromosome by homologous recombination, then only two segregated ascospores are viable, namely those which carry the essential gene. The two remaining segregated ascospores are not viable because they lack the essential gene.

In order to check if the genes studied by this method are really essential or if the homologous recombination leads to an alteration of an essential gene adjacent to the gene locus of the gene studied, the heterozygote diploid S. cerevisiae mutant strain is transformed with a centromere plasmid containing said studied gene.

A tetrad analysis is carried out on the transformants. When four instead of two viable segregates are obtained, then the studied gene contained in the centromere plasmid can complement the defect of the two non-viable haploid S. cerevisiae cells/mutant strains, which demonstrates that the studied S. cerevisiae gene is essential.

Preferably, plasmids present in low copy number, e. g. one or two copies per cell are used as centromere plasmids.For example plasmids pRS313, pRS314, pRS315 and pRS316 (Sijkorski, R. S. and Hieter, P. (1989) Genetics 122: 19- 27) or similar plasmids can be used for this purpose. Preferably, the studied genes are integrated in said plasmids including their 3'-and 5'-end non-coding regions.

Individual S. cerevisiae genes may be studied using the above-described method, their sequences being totally or partially known. The complete genomic sequence of S. cerevisiae was made accessible to the public via the WWW (World Wide Web) on April 24,1996.

Different possibilities exist to have access to the S. cerevisiae genomic DNA sequence via the WWW.

MIPS (Munich information Centre of Protein Sequence) Address http://speedY. mips. biochem. mpa. de/mips/yeast/

SGD (Saccharomyces Genome Database, Stanford) Address http://genome-www. stanford. edu/Saccharomyces YPD (Yeast Protein Database, Cold Spring Harbor) Address http://www. proteome. com/YPDhome. html

The complete S. cerevisiae DNA sequence is also accessible via FTP (file transfer protocol) in Europe (e. g. at the address: ftp. mips. embnet. org) in the U.S.A. (address: genome-ftp. stanford. edu) or in Japan (address: ftp. nig. ac. jp).

7 essential genomic S. cerevisiae genes have been identified by this way: YDR325w, YJL039c, YOR110w, YNL256w, YBR102c, YIR012w and YMR212c

The essential genes of S. cerevisiae are then used to identify corresponding functionally similar genes in other mycetes.

By functionally similar genes in other mycete species, is meant genes which have a function similar or identical to that of the identified essential genes of S. cerevisiae. Functionally similar genes in other mycetes may, but need not be homologous in sequence to the corresponding essential S. cerevisiae genes. Functionally similar genes in other mycetes may exhibit only moderate sequence homology at the nucleotide level to the corresponding essential S. cerevisiae genes. By moderate sequence homology it is meant in the present invention genes having a sequence identity, at the nucleotide level, of at least 50%, more preferably of at least 60% and most preferably of at least 70%.

In addition, functionally similar genes in other mycetes may, but need not encode proteins homologous in sequence to the proteins encoded by the essential S. cerevisiae genes. Functionally similar proteins in other mycetes may exhibit moderate protein sequence homology to the proteins encoded by the essential S. cerevisiae genes.

By moderate protein sequence homology is meant in the present invention proteins having a sequence identity, at the amino-acid level, of a least 40%, preferably of at least 50%, more preferably of at least 60% and most preferably of at least 70%.

Genes homologous in sequence may be isolated according to known methods, for example via homologous screening (Sambrook, J. et al. (1989) Molecular Cloning Cold Spring Harbor Laboratory Press, N. Y.) or via the PCR technique using specific primers from genomic libraries and/or cDNA libraries of the corresponding mycetes.

According to one embodiment, genes homologous in sequences are isolated from cDNA libraries. In order to find out functionally similar genes in other mycetes, mRNA is isolated from mycete species to be studied according to known methods (Sambrock et al. 1989) and cDNA is synthesized according to known methods (Sambrock et al.1989; or cDNA synthesis kits, e. g. from STRATAGENE).

The prepared cDNA is directionally integrated in a suitable expression vector.

For example, synthesis of the first cDNA strand may be carried out in the presence of primers having appropriate restriction sites in order to allow a subsequent cloning in the proper orientation with respect to the expression vector promoter. As restriction sites, any known restriction site may be used. As a primer, for instance the following primer, 50 nucleotides long may be used:5'GAGAGAGAGAGAGAGAGAGAACTAGTXXXXXXTTTTTTTTTTTTTTTTT-3' (SEQ ID NO: 20)

The sequence (X) 6 represents an appropriate restriction site, for example for XhoI.

After two-strand synthesis, the cohesive ends of the double stranded cDNA are filled (blunt end) and the cDNA ends are then ligated using a suitable DNA adaptor sequence. The DNA adaptor sequence should contain a

restriction site which should be different from the restriction site used in the primer for the synthesis of the first cDNA strand. The DNA adaptor may comprise for example complementary 9-or 13-mer oligonucleotides, whose ends represent the cohesive end of a restriction site. These ends may be for example a EcoRI-site:

```
5'  XXXXXGGCACGAG  3'
3'      XCCGTGCTC  5'
```

The single-stranded X in the adaptor sequence represent the cohesive end of a restriction site.

The cDNA provided with corresponding adaptor sequences is then cleaved using restriction endonuclease, whose recognition site was used in the primer for the synthesis of the first cDNA strand, for example XhoI. The cDNA thus obtained would have according to this example 3'-XhoI and 5'-EcoRI protruding ends and could be therefore directionally integrated into an expression vector cleaved with XhoI and EcoRI.

As expression vectors, among others, E. coli/S. cerevisiae shuttle vectors, i. e. vectors usable in E. coli as well as in S. cerevisiae are suitable. Such vectors may then be amplified for instance in E. coli. As expression vectors, those which are present in a high copy number as well as those present in a low copy number in S. cerevisiae cells can be used. For this purpose, for example vectors selected in the group consisting of pRS423 -pRS426 (pRS423, pRS424, pRS425, pRS426) and/or pRS313- pRS316 (pRS313, pRS314, pRS315, pRS316) (Sikorki, R. S. and Hieter, P. (1989) Genetics 122: 19-27; Christianson T. W. et al. (1992) Gene 110: 119-122) are suitable.

Expression vectors should contain appropriate S. cerevisiae promoters and terminators. In case they do not have these elements, the corresponding promoters and terminators are inserted in such a way that a subsequent incorporation of the generated cDNA remains possible.

Particularly suitable are the promoters of S. cerevisiae genes MET25, PGK1, TPI1, TDH3, ADHI, URA3. One may use promoters of the wild-type gene in non modified form as well as promoters which were modified in such a way that certain activator sequences and/or repressor sequences were eliminated. As terminators, for example the terminators of the S. cerevisiae genes MET25, PGK1, TPI1, TDH3, ADHI, URA3 are suitable.

According to another embodiment, genes homologous in sequence are isolated from genomic libraries. Genomic DNA libraries from mycetes can be prepared according to procedures known (for example as described in Current Protocols in Molecular Biology, John Wiley and Sons, Inc).For example, genomic DNA from mycetes can be prepared using known methods for yeast cell lysis and isolation of genomic DNA (for example commercially available kits from Bio101, Inc). The genomic DNA can be partially digested using a restriction enzyme such as Sau3AI and the fragments are size-selected by agarose gel electrophoresis. DNA fragments having for example a size of 5-10kb are then purified by classical methods (as for example, using Gene Clean kit from Bio101) and inserted in a E. coli/yeast shuttle vector such as YEP24 (described e. g. by Sanglard D., Kuchler K., Ischer F., Pagani J-L., Monod M. and Bille J., Antimicrobial Agents and Chemotherapy, (1995) Vol. 39 No11, P2378-2386) cut by a restriction enzyme giving compatible ends (for example BamHI for Sau3AI-cut genomic DNA). The resulting expression library can be amplified in E. coli. However any known method, appropriate for the preparation of a genomic library, can be used in the present invention.

In order to find the genes in the studied mycete species, which are functionally similar to essential genes of S. cerevisiae, one S. cerevisiae essential gene is placed under control of a regulated promoter, either as an integrative (1) or extrachromosomal (2) gene.

1. For the integration of a regulated promoter in the S. cerevisiae genome, one replaces the native promoter of the selected essential gene by the regulated promoter, for example by homologous recombination via PCR (Guldener et al. (1996). The homologous recombination via PCR can be carried out for example in the diploid S. cerevisiae strain CEN. PK2. The successfull integration into one chromosome can be checked in haploid cells following tetrad analysis.

Using the tetrad analysis, one obtains four viable ascospores, wherein in two haploid segregates, the selected essential gene is placed under the control of the native promoter, while the essential gene in the two remaining segregates is placed under the control of the regulated promoter.

The last mentioned haploid segregates are used for the transformation with the cDNA or the genomic DNA present in the recombinant vector.

2. Using the extrachromosomal variant, the selected essential S. cerevisiae gene is first inserted in a suitable expression vector, for example a E. coli/ S. cerevisiae shuttle vector. For this purpose, the essential gene may be amplified via PCR from genomic S. cerevisiae DNA starting from the ATG initiation codon up to and including the termination codon. The primers used for this purpose may be constructed in such a way that they contain recognition sites for appropriate restriction enzymes, facilitating a subsequent insertion under control of a regulated promoter in an expression vector.

The recombinant expression vector with the plasmid copy of the essential S. cerevisiae gene under the control of a regulated promoter is subsequently used for the transcomplementation of the corresponding mutant allele. The corresponding mutant allele may be selected from the heterozygote-diploid mutant strains prepared by eliminating, partially or totally, by homologous

recombination an essential mycete gene listed above and as described above.

The expression vector with the selected essential S. cerevisiae gene is transformed in the corresponding heterozygote-diploid mutant strain carrying instead of the selected essential S. cerevisiae gene, a selection marker gene. The transformants are isolated by selection based on the auxotrophic marker contained in the expression vector used. The thus transformed heterozygote-diploid mutant strains are submitted to a tetrad analysis. One obtains four viable segregates. By retracing the corresponding markers of the mutant allele and the expression vector, the transformed wild-type segregates may be distinguished from segregates which do not contain the genomic copy of the essential gene. Segregates, which do not contain the genomic copy of the selected essential gene, are designated as trans-complemented haploid mutant strains. They are subsequently used for transformation with cDNA or genomic DNA libraries from other mycete species present in appropriate vectors.

As regulated promoters, inducible or repressible promoters may be used. These promoters can consist of naturally and/or artificially disposed promoter sequences.

As regulated promoters, for example the promoters of GAL1 gene and the corresponding promoter derivatives, such as for example promoters, whose different UAS (upstream activation sequence) elements have been eliminated (GALS, GALL; Mumberg, J. et al. (1994) Nucleic Acids Research 22: 5767-5768) may be used. As regulated promoters, promoters of gluconeogenic genes may also be used, such as e. g. FBP1, PCK1, ICL1 or parts therefrom, such as e. g. their activation sequence (UAS1 and/or UAS2) or repression sequence (URS, upstream repression sequence) (Niederacher et al. (1992), Curr. Genet. 22: 636-670; Proft et al.(1995) Mol. Gen. Gent. 246: 367-373; Schüller et al. (1992) EMBO J; 11: 107-114; Guarente et al. (1984) Cell 36: 503- 511).

A S. cerevisiae mutant strain modified in this manner can be cultivated under growth conditions, in which the regulated promoter is active, so that the essential S. cerevisiae gene is expressed. The S. cerevisiae cells are then transformed with a representative quantity of the library containing the studied mycete species cDNA or genomic DNA. Transformants express additionally the protein whose coding sequence is present in the recombinant vector.

The method contemplates that the growth conditions may be modified in such a way as to inhibit the regulated promoter, under the control of which is the selected essential gene. Especially, growth conditions may be changed by replacing the growth medium. When for example the GAL1 promoter or a derivate thereof is used, one can replace the galactose-containing medium (induced state) by a glucose-containing medium (repressed state).

These modified conditions are lethal for the S. cerevisiae cells in which the recombinant vector does not carry the functionally similar genomic DNA or cDNA of the studied mycete species. On the contrary, the S. cerevisiae cells in which the recombinant vector expresses a functionally similar coding sequence of the studied mycete species, are viable, since in these cells the lethal metabolic defect is complemented by the protein encoded by the functionally similar gene.

The method contemplates that the recombinant vector (the plasmid) is isolated from the surviving transformants using known method (Strathern, J. N. and Higgins, D. R.(1991). Plasmids are recovered from yeast into Escherichia coli shuttle vectors in: Guthrie, C. and Fink, G. R. Methods in Enzymology, Volume 194. Guide to yeast genetic and molecular Biology. Academic Press, San Diego, 319-329) and the cDNA or genomic DNA is analyzed using DNA-analysis methods such as DNA sequencing. (Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74: 5463-5467)

Essential S. cerevisiae genes may thus be used for the identification of functionally similar genes and/or genes homologous in sequence in other mycetes, especially essential genes functionally similar and/or homologous in sequence in mycetes pathogenic to human, animal and plants.For this purpose for example mycetes of the classes phycomycetes or eumycetes may be used, in particular the subclasses basidiomycetes, ascomycetes, especially mehiascomycetales (yeast) and plectascales (mould fungus) and gymnascales (skin and hair fungus) or of the class of hyphomycetes, in particular the subclasses conidiosporales (skin fungus) and thallosporales (budding or gemmiparous fungus), among which particularly the species mucor, rhizopus, coccidioides, paracoccidioides (blastomyces brasiliensis), endomyces (blastomyces), aspergillus, penicilium (scopulariopsis), trichophyton (ctenomyces), epidermophton, microsporon, piedraia, hormodendron, phialophora, sporotrichon, cryptococcus, candida, geotrichum and trichosporon.

Of particular interest is the use of Candida Spp. especially Candida albicans, Candida glabrata, Aspergillus Spp., especially Aspergillus fumigatus, Coccidioides immitis, Cryptococcus neoformans, Histoplasma capsulatum, Blastomyces dermatitidis, Paracoccidioides brasiliens and Sporothrix schenckii.

Starting from the genes of S. cerevisiae, identified according to the above-described method, Applicants cloned corresponding essential genes from C. albicans i. e. CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, by the following method.

First, oligonucleotide (s) is (are) selected in the sequence of the S. cerevisiae gene or a homologous C. albicans sequence in order to amplify the corresponding fragment of C. albicans. After cloning, the obtained fragment (exhibiting a sequence of about several hundred bp) is used as a probe for screening a C. albicans

(genomic) DNA library. The screening may include the following steps: clones were spread on dishes, covered with filters where the DNA was crosslinked to the filters, filters are hybridized, the positive colonies are then detected. The selected clone (s) is (are) then sequenced.

The method contemplates that essential mycete genes are used to identify substances which may inhibit partially or totally the functional expression of these essential genes and/or the functional activity of the encoded proteins. Substances may be identified in this fashion, which inhibit mycetes growth and which can be used as antimycotics, for example in the preparation of drugs.

The present invention especially covers a method for screening such inhibiting substances wherein an essential gene from C. albicans selected from CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, or a functionally similar gene in another pathogenic mycete or the corresponding encoded protein is used as target.

By functionally similar genes in other pathogenic mycete species, is meant genes which have a function similar or identical to that of the identified essential genes of C. albicans. Functionally similar genes in other pathogenic mycetes may, but need not be homologous in sequence to the corresponding essential C. albicans genes. Functionally similar genes in other pathogenic mycetes may exhibit only moderate sequence homology at the nucleotide level to the corresponding essential C. albicans genes. By moderate sequence homology it is meant in the present invention genes having a sequence identity, at the nucleotide level, of at least 50%, more preferably of at least 60% and most preferably of at least 70%.

In addition, functionally similar genes in other pathogenic mycetes may, but need not encode proteins homologous in sequence to the proteins encoded by the essential C. albicans genes. Functionally similar proteins in other mycetes may exhibit moderate protein sequence

homology to the proteins encoded by the essential C. albicans genes.

By moderate protein sequence homology is meant in the present invention proteins having a sequence identity, at the amino-acid level, of a least 40%, preferably of at least 50%, more preferably of at least 60% and most preferably of at least 70%.

A particular feature of this method is that essential mycete genes or the corresponding encoded proteins are used as targets for the screening of the substances. The method contemplates that essential C. albicans genes as well as functionally similar genes and/or genes homologous in sequence of other pathogenic mycetes or the corresponding encoded proteins may be used as targets.

According to one embodiment of the screening method of the invention, mycetes cells are provided, which contain the essential gene used as target, and those cells are incubated with the substance to be tested. By this way, the growth inhibitory effect of this substance with respect to the essential target gene is determined.

The mycetes cells which express the essential target gene to a different degree are used, and these cells are then incubated with the substance to be tested and the growth inhibitory effect of this substance is determined.

The method includes the use of two or more mycetes cells, or strains derived therefrom, which differ in that they express the essential target gene to a different degree.

For example, two, three, four, five, ten or more mycetes cells or the corresponding mycetes strains may be comparatively analysed with respect to the growth inhibitory effect of a substance used in a defined concentration. Through such expression degree series, antimycotic substances may be distinguished from cytotoxic or inactive substances.

A particular embodiment of the method includes the use of haploid mycetes cells/strains for the screening, especially haploid S. cerevisiae cells/strains.

The method contemplates the integration of the essential gene selected as a target in a suitable expression vector.

As expression vectors E. coli/S. cerevisiae shuttle vectors are for example suitable. Especially vectors differing in their copy number per cell may be used. Therefore, one may use vectors, which are present in the transformed S. cerevisiae cells in a high copy number, or one can also use those with a low copy number. One embodiment comprises the use of expression vectors which allow the integration of the target gene in the S. cerevisiae genome.

For example the vectors pRS423, pRS424, pRS425, pRS426, pRS313, pRS314, pRS315, pRS316, pRS303, pRS304, pRS305, pRS306 (Sikorki and Hieter, 1989; Christianson et al. 1992) are appropriate as expression vectors.

The vectors of the series pRS423-pRS426 are present in a high copy number, about 50-100 copies/cell. On the contrary, the vectors of the series pRS313-pRS316 are present in a low copy number (1-2 copies/cell). When expression vectors from these two series are used, then the target gene is present as an extrachromosomal copy. Using the vector of the series pRS303-pRS306 allows the integration of the target genes into the genome. Using these three different expression vector types allows a gradual expression of the studied functionally similar essential gene.

The method includes that the growth inhibitory effect of substances with respect to mycetes cells/strains is comparatively determined using expression vectors differing for instance in the copy number of the vector/cell.

Such cells may express the essential target gene to a different degree and may exhibit a graduated reaction with respect to the substance.

The method includes also, that a target gene expression of different strength is obtained in different mycetes cells (regulated overexpression) by insertion of the target gene in the expression vector between specific selected S. cerevisiae promoters and terminators. S. cerevisiae promoters which are constitutively expressed, but with different strength, are suitable. Examples for such promoters are native promoters of S. cerevisiae genes MET25, PGK1, TPI1, TDH3, ADH1, URA3, TRP1, as well as corresponding derivatives therefrom, for example promoter derivatives without specific activator and/or repressor sequences.

Regulated promoters are also appropriate for the graduated over-expression of the target gene. The native promoters of the GAL1 genes and/or corresponding derivates thereof, for example promoters, in which different UAS elements have been eliminated. (GALS, GALL; Mumberg et al.(1994) Nucleic Acids Research 22: 5767-5768) as well as promoters of gluconeogenic genes, for example the promoters FBP1, PCK1, ICL1, or parts thereof, for example their activator- (UAS1 or UAS2) or repressor- (URS) sequences are used in corresponding non activable and/or non repressible test promoters (Schüller et al. (1992) EMBO J. 11: 107-114) Guarente et al. (1984) Cell 36: 503-511; Niederacher et al.(1992) Curr. Genet. 22: 363-370; Proft et al. (1995) Mol.Gen. Genet. 246: 367-373).

In the expression vector terminator for example the terminator sequence of S. cerevisiae genes MET25, PGK1, TPI1, TDH3, ADHI, URA3 may be used.

The method includes that by the use of cleverly selected expression vector types and/or the preparation of suitable expression vectors, eventually using promoters of different strength and differently regulated promoters, a

series of expression vectors may be constructed, all containing the same target gene, but differing in that they express the target gene to a different extent.

The method includes the transformation of the expression vector in haploid wild-type cells of S. cerevisiae. The thus obtained S. cerevisiae cells/strains are cultivated in liquid medium and incubated in the presence of different concentrations of the tested substance and the effect of this substance on the growth behaviour of the cells/strains expressing the target gene to a different degree is comparatively analysed. The method also includes that haploid S. cerevisiae cells/strains, transformed using the respective expression vector type without target gene, are used as a reference.

The method includes that the screening of the substances can be carried out in different media using regulated promoters, especially GAL1 promoter and its derivates (GALS and GALL), since the expression degree may be largely influenced by the choice of the respective medium. Thus, the expression degree of the GAL1 promoter decreases in the following fashion: 2 % galactose > 1 % galactose + 1 % glucose > 2 % glycerine > 2 % glucose.

The effect of the substances inhibiting the growth of wild-type cells of S. cerevisiae, may be partially or totally compensated by the overexpression of the functionally similar gene of another mycete species.

According to one embodiment, the method for screening antimycotic substances is carried out in vitro by contact of an essential or functionally similar gene or the corresponding encoded protein with the substance to be tested and determination of the effect of the substance on the target. Any in vitro test appropriate for determining the interaction of two molecules, such as a hybridization test or a functional test, can be used (e. g. enzymatic tests which are described in details in Bergmeyer H. U., Methods of Enzymatic Analysis, VCH Publishers). If the

screening is carried out using the encoded protein as the target, then the corresponding essential gene is inserted by any suitable method known in the art, such as PCR amplification using a set of primers containing appropriate restriction sites, (Current Protocol in Molecular Biology, John Wiley and Sons, Inc) into an expression system, such as E. coli, Baculovirus, or yeast, and the expressed protein is then completely or partially purified by a method known in the art. Any purification method appropriate for the purification of expressed proteins, such as affinity chromatography can be used. If the target protein function is known, a functional test can then be carried out in which the effect of the antimycotic substance on the protein function is determined. If the protein function is unknown, substances which can interact with the target protein, e. g. which bind to the encoded protein, can be tested. In such a case a test such as protection of the target protein from enzymatic digestion by appropriate enzymes can be used.

According to one specific embodiment, the method for screening antimycotic substances corresponds to an enzymatic assay wherein the activity of dihydropteroate synthase (DHPS) and/or 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK) is determined; the enzymatic essay can be such as diclosed in"Bergmeyer H. U., Methods in Enzymatic analysis, VCH Publishers".

Dihydropteroate synthetase (DHPS) catalyses the condensation of 6-hydroxymethyl-7,8-dihydropterin pyrophosphate to para-aminobenzoic acid to form 7,8-dihydropteroate which corresponds to the second step in the three-step pathway leading from 6-hydroxymethyl-7,8-dihydropterin to 7,8-dihydrofolate. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK) catalyzes the attachment of pyrophosphate to 6-hydroxymethyl-7,8-dihydropterin to form to 6-hydroxymethyl-7,8-dihydropterin pyrophosphate which corresponds to the first step in a

three-step pathway leading to 7,8-dihydrofolate. All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial agents. Consequently, these enzyme activities are essential to the microorganisms, and are absent in man.

The method also includes the identification of genes which are functionally similar and/or homologous in sequence to essential C. albicans genes from humans, animals or plants. The corresponding human, animal or plant genes may optionally be used as target genes in the method in order to test if antimycotic substances exhibit an effect on these target genes.

A particular advantage of the method is that in this way substances may be identified which efficiently inhibit mycetes growth and also the influence of these substances on corresponding functionally similar genes and/or genes homologous in sequence to essential C. albicans genes from human, animal or plants may be determined.

The method includes also the possibility to check the existence of functionally similar genes and/or human, animal or plant genes homologous in sequence to the corresponding essential mycete genes, for example by checking homology of the identified essential mycete genes or parts thereof with human, animal or plant sequence genes available in data banks. In this way, it is possible to select at an early stage from the identified essential mycete genes, depending on the aim, those for which no functionally similar gene and/or no human gene homologous in sequence exist, for example.

Thereby, the method offers a plurality of possibilities to identify selectively substances with

antimycotic effects, with no harmful effect on human beings.

For example, it is possible to identify substances usable for the preparation of drugs for the treatment of mycosis or prophylaxis in immunodepression states. These substances may be used for example for the manufacture of drugs usable for the treatment of mycotic infections, which occur during diseases like AIDS or Diabetes. Substances which may be used for the fabrication of fungicides, especially of fungicides which are harmless for humans and animals, can also be identified.

Furthermore, the method offers the possibility to identify antimycotic substances, which selectively inhibit growth of specific mycete species only.

The screening method is particularly advantageous inasmuch as it is sufficient to know whether the genes are essential, one does not need any additional information regarding the function of the essential genes or the function of the encoded proteins. In addition, it is particularly advantageous for the identification of functionally similar genes to essential S. cerevisiae gene, in other mycetes where the DNA sequence is not available for many of these genes.

According to another aspect the invention provides an antibody directed against the protein encoded by the CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039, gene or a polypeptidic fragment thereof. The term "antibody"encompasses monoclonal and polyclonal antibodies. Said antibodies can be prepared by method well known in the art such as those disclosed in"Antibodies, a laboratory manual", Ed. Harbow and David Lane. Cold Spring Harbor Laboratory Eds., 1988.

According to another aspect the present invention provides a kit for the diagnosis of fungal infections comprising a gene selected from the group consisting of CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 and

CaJL039, a functionally similar gene thereof, a functional fragment thereof, the corresponding encoded protein, a functional polypeptide fragment thereof or an antiboby directed against the protein encoded by CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 or CaJL039 gene or by a functionally similar gene, or a polypeptidic fragment thereof. Such kits can be prepared using any suitable method well known in the art.

Examples

Example 1: CaNL256

The Internet site of Standford (http://candida.standford.edu/) gives access to preliminary sequences of the genome of C. albicans. One of these sequences has homology with the YNL256 gene of S. cerevisiae. Two oligonucleotides were selected in this sequence (5'-ATTCATCCCATCAGTGCAGAAAG-3' (SEQ ID NO:27) and 5'- ATTGACCAATAGCTCTAATTAATG-3') (SEQ ID NO:28) in order to amplify the corresponding fragment of C. albicans. After cloning, we obtained a sequence of 399 bp close to the expected sequence (SEQ ID NO: 1). The deduced protein was compared with the one of YNL256, evidencing 53% similarity and 43% identity (fig. 1). This fragment of 399 bp of C. albicans was used as a probe for screening a genomic library of C. albicans. The latter was prepared by partial digestion of genomic DNA of C. albicans by Sau3AI and cloning into the YEP24 vector at the BamHI site. The clones of the library were then spread at a density of 2000 clones per dish. Each dish was covered by a nitrocellulose filter which was then successively treated with: NaOH, 0.5M, 5 minutes; Tris, 1M, pH 7.7,5 minutes; Tris, 0.5M, pH 7.7, Nacl, 1.25M, 5 minutes. After drying, the filters were kept for 2 hours at 80°C. Prehybridization and hybridization were carried out in a buffer of 40% formamide, 5xSSC, 20 mM Tris pH 7.7 1xDenhardt 0.1% SDS. The probe was labeled with 32P with the RediPrime kit and dCTP from Amersham

UK.Hybridization took place over 17 hours at 42°C. The filters were then washed in 1x SSC, 0.1% SDS, three times for 5 minutes at room temperature and then twice for 30 minutes at 60°C, and were then submitted to autoradiography overnight. The colonies corresponding to the spots obtained were reisolated by re-spreading at low density followed by further hybridization. Three clones were thus obtained (out of 40,000), which were sequenced on an ABI 377 apparatus. The sequences were compiled using the ABI software and then analysed using the GCG software package.One of these three clones turned out to contain the complete coding sequence corresponding to the probe used; this gene was called CaNL256, whose sequence is represented in SEQ ID NO: 2. CaNL256 has 52% of nucleotides identical to YNL256 of S. cerevisiae. The coding region is shorter at the N-terminus. For translation to amino acids, account was taken of the fact that, in C. albicans, the CTG codon is translated to Serine (there are 3 CTG codons in CaNL256). The deduced protein (SEQ ID NO: 3) had 40% amino acids identical with YNL256 of S. cerevisiae and 41% with FAS (folic acid synthase) of Pneumocystis carinii.Investigation into the databases using the Blast software showed homology of two parts of the CaNL256 protein with, respectively, the bacterial enzymes Dihydropteroate Synthase (EC 2.5.1.15) (DHPS) of Haemophilus influenzae, Staphylococcus haemolyticus, Neisseria meningitidis, Streptococcus pneumoniae, Bacillus subtilis, Clostridium acetobutylicum, Escherichia coli, Mycobacterium leprae (P value less than e-28) and 7,8-dihydro-6- hydroxymethylpterin-pyrophosphokinase (EC 2.7.6.3) (HPPK) of Bacillus subtilis, Escherichia coli, Haemophilus influenzae, Streptococcus pneumoniae (P value less than e- 20). The units characteristic of DHPS and HPPK are also found in CaNL256.

Example 2: CaBR102

The Internet site of Stanford

(http://candida. stanford. edu/) give access to preliminary sequences of the genome of C. albicans. One of these sequences has homology with the YBR102 gene of S. cerevisiae. Two oligonucleotides were selected in this sequence (5'-AGTATTCAATTGGGTATTCC-3'(SEQ ID NO:29)and 5/-CCGGCATCATCAGTAACTCC-3')(SEQ ID NO:30) in order to amplified the corresponding fragment of C. albicans. After cloning, we obtained a sequence of 647 bp (SEQ ID NO: 4~~3~~). The deduced protein was compared with the one of YNL102, evidencing 35% similarity and 26% identity (fig. 2). This fragment was amplified using Pfu polymerase (Stratagene). The PCR product was purified (High Pure PCR Product Purification Kit, Boehringer Mannheim) and used as a probe for screening a C. albicans genomic DNA library. The latter was prepared by partial digestion of C. albicans genomic DNA with SauIIIA and cloning into the YEP-24TRP1 vector at the BamHI restriction site. 40,000 clones of the library were then spread at a density of 2000 clones per dish. Each dish was covered by a nitrocellulose filter (Membrane Hybond N', Amersham) which was then successively treated with : 1.5 M NaCl/0.5 M NaOH, 5 minutes; 1.5 M NaCl/0.5 M Tris-HCl pH 7.2/1 mM EDTA, 3 minutes, twice; DNA was crosslinked to the filters (Amersham Life Science, ultra violet crosslinker). The probe (100 ng) was labelled with 32p using the RediPrime kit and dCTP (Amersham Life Science). The filters were hybridized in a buffer containing 30% formamide, 5 x SSC, 5% Denhart's solution, 1% SDS, 100 yg/ml salmon sperm DNA and a probe concentration of 106 cpm/ml at 42°C for 16 h. The membranes were then washed three times at room temperature in 2 x SSC/0.1% SDS for 5 minutes each and three times in 1 x SSC/0.1% SDS at 60°C for 20 minutes each. the filters were then exposed overnight to an X-ray film. The colonies corresponding to the positives clones were isolated and screened a second time by the same procedure. Two positives clones were finally obtained, which were sequenced on an

ABI377 apparatus. the sequences were compiled using ABI software and then analysed using the GCG software package. The nucleotide sequences of these two clones were identical and contained the complete coding sequence corresponding to the probe used, this gene was called CaBR102, whose sequence is represented in SEQ ID No: 5 and 6~~4~~. The translation of this nucleotide sequence was examined, account was taken of the fact that in C. albicans the CTG codon is translated to serine (there are 3 CTG codons in CaBR102). The deduced protein has 24% identity to S. cerevisiae gene YBR102.

Example 3: CaIR012

Chromosomal DNA from the C. albicans strain Caf2-1 was isolated using Yeast Cell Lysis prep Kit and Genome DNA Kit from BI0101. A 343 bp fragment from C. albicans genomic DNA (SEQ ID NO: 7~~5~~) was amplified with the oligonucleotide primers CaYIR012-5' (5'-GACGTCGTAGACGATACTCAAGAAG-3') (SEQ ID NO:31) and CaYIR012-3' (5'-CTGCAGTAAACCCTCCAGATATAACAG-3') (SEQ ID NO:32) by PowerScript DNA polymerase (PAN Systems GmbH) using the hot start technique. The PCR product was purified from the agarose gel and labeled with fluorescein (Gene image random prime labelling module, Amersham Life Science) according to the manufacturer's instructions. Plasmid DNA from E. coli was isolated using Qiagen columns as recommended by the manufacturer. Screening the kZAPII C. albicans cDNA library was performed following the manufacturer's instructions (Stratagene Ltd.). Nylon filters (Schleicher&Schuell) were lifted from LB-plates (150 mm) with 15000 pfu/plate, denatured 5 min in 1.5M NaCL, 0.5M NaOH, neutralized 3 min in 1.5M NaCl, 0.5M Tris-HCl pH8.0, washed 3 min in 0.2 M Tris-HCl pH 7.5,2xSSC and DNA was crosslinked to the filters (Stratagene W crosslinker). The filters were prehybridized 4 h at 60 °C and hybridized with the fluorescein-labeled DNA probe overnight at 60 °C. Detection was performed with Anti-fluorescein AP conjugate (Signal amplification module for

the FluorImager, Amersham LIFE SCIENCE) and analysed after 20 h with a Fluorimager (Storm 860, Molecular Dynamics). Positive plaques were picked and incubated with 0.5 ml SM-buffer (100mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH7.5,0,01% gelatin). The selected clones were diluted, titered with host cells XL1- Blue and screened and purified a second time by the same procedure. Finally, the pBluescript SK (-) phagemid containing the DNA insert of interest was rescued by the ExAssist Helper Phage system according to the Stratagene protocol. From a total of 75000 screened plaques, 3 positive clones were identified. pBluescript SK (-) phagemid DNA was isolated, sequenced with T3 and T7 primers and the sequences were extended with custom-synthesized oligonucleotide primers. Nucleotide sequence analyses were performed with the Gene Data software package (Gene Data AG, Basel Switzerland). Similarity searches with the Swissprot database were conducted with the BLAST program (Gish, Warren and David J. States (1993). Identification of protein coding regions by database similarity search.Nat. Genet. 3: 266-72.). One of these three clones turned out to contain the complete coding sequence corresponding to the probe used; this gene was called CaIR1012, whose sequence is represented in SEQ ID NO: 8 and 9~~6~~.

Example 4: CaJL039

The CaJL039 sequence is depicted in SEQ ID No 10 and 11~~7~~.

The CaJL039 gene was cloned based on gene fragment data issued from the public Stanford Candida albicans sequencing database.

(a) A fragment that showed homology to Saccharomyces cerevisiae YJL039c was identified, the sequence of which is given in SEQ ID No 12~~8~~.

Using the procedure disclosed in example 3 with the oligonucleotide primer pair (Ca039s: TAG CTC AAC CTA CCA CCA ATC (SEQ ID NO:33)/Ca039r: ATC ACA AGA CTG TCA ATG TAA

AT)(SEQ ID NO:34), a short PCR fragment (234 base pairs long) was amplified for screening a Candida albicans cDNA lambda ZAP II library (gift of Alistair Brown, Aberdeen).

Three positive clones of the 3 coding region were obtained. [(#]21t7, 11t3,21t3).

(b) 3-and 5-extension of the internal fragment using the primer walking method

The Sanglard genomic Candida DNA library with the YEp24 vector backbone was used for further amplification of 3-and 5-coding sequences. Amplification was carried out by using the following vector-specific oligonucleotide primers and CaJL039 fragment-specific primers:

Cggaattcctatcgactacgcgatcatgg(SEQ ID NO:35): YEp24for (vector specific)

Gcgaattccgatataggcgccagcaac(SEQ ID NO:36): YEp24ba (vector specific)

Caattgctttgactcgggtgttattaagt(SEQE ID NO:37): Ca039-51 (CaJL039: 5fishing)

Tcttggcacaacttgataagaatctgt(SEQ ID NO:38): C1039-52

Taggtgtacgcgaaagccaagtagaac(SEQ ID NO:39): Ca039-53 ()

Ttgttaatcgtacacctaaggtgttgac(SEQ ID NO:40): Ca039-31 (CaJL039: 3fishing)

Ttgcagattgatgctagcaatgtatttg(SEQ ID NO:41): Ca039-32

Using the technique of primer walking, the complete 5'-sequence could be amplified (clone 14b-1-1 and clone 17b-3-4).

The missing 3-sequence was available from GTC PathoGenome Release 5.0, contig #2830.

An interacting protein (C82, component for RNA polymerase III in yeast) has been identified.

Example 5: CaOR110

5.1. CaOR11O

The CaOR110 sequence is depicted in SEQ ID No 13 and 14~~9~~.

CaOR110 was cloned based on gene fragment data issued from the public Stanford Candida albicans sequencing database.

(a) A small ScOR110-homologous fragment was used in a hybridization experiment to identify CaOR110 clones in a Candida Albicans lambda ZAPII cDNA library (from Alistair Brown). Alignement of Candida Albicans CaOR110 sequence with the fragment used for hybridization is given in figure 3. The homologous fragment sequence is given in SEQ ID No.25~~17~~.

(b) 3'-and 5'-Extension of the internal fragment: The Sanglard genomic Candida DNA library (received from RMV) in the YEP24 vector backbone was used for the amplification of 3'-and 5'-coding and non-coding sequences. This amplification was done by using the vector- specific oligos (directional towards the insert) and CaOR110 fragment-specific oligos (directional towards the vector flanking sequences) described below:

cggaattcctatcgactacgcgatcatgg(SEQ ID NO:35) : YEP24for

gcgaattccgatataggcgccagcaac(SEQ ID NO:36) : YEP24ba

cgggatccggtaaccaattggatctataaccgtg(SEQ ID NO:42) : 110-ba-150

gcggatcctggtgcccttggtggtgaatg(SEQ ID NO:43) : CaYOR110A

gcggatccctcacaatatgacgattgaaact(SEQ ID NO:44) : CaYOR110B

ggcgtcgactcaggcgccagttttacgtacttcaaattcatc(SEQ ID NO:45) : CaYOR110C

tgtgaattcttgacacagggtga(SEQ ID NO:46) : CaYOR110D

caaaccttcagcacaactcca(SEQ ID NO:47) : CaYOR110E

The finally assembled sequence that included also 3'- and 5'-non-coding sequences was verified by sequencing. The coding region was subcloned into the p414RSGALL-vector.

The map is depicted in Fig. 4.

The homologous yeast ORF (YOR110w) has been described as the transcription factor subunit TFC7 interacting with

TFC1 in the TFIIIC polymerase complex (Manaud et al., 1998, Mol. Cell. Biol. 18; 3191-3200).

5.2. CaOR110 splice variant

For CaOR110, an additional splice variant was identified. The clones for the splice variant of CaOR110 were obtained from a Candida albicans cDNA library.

The sequence is depicted in SEQ ID No. 15 and 16~~10~~.

The splice variant uses the donor site"gtacgt"at position 907 of the original CaOR110 sequence. Acceptor site is at 1047. The map is disclosed in Fig. 5.

The alignement of the original CaOR110 and the splice variant is given in fig. 6.

Example 6: CaMR212

The CaMR212 sequence is depicted in SEQ ID No. 17 and 18~~11~~

(a) CaMR212 was cloned based on gene fragment data from the public Standford Candida albicans sequencing database.

The sequence of a fragment showing homology (Blast search) to the Saccharomyces cerevisiae gene YMR212c is given in SEQ ID 19~~12~~.

Based on these data, the following oligos were designed that allow amplification of this fragment (490 bp-fragment) from genomic Candida albicans DNA. Oligos:

CaYMR212for: 5'-cacctgtgaacaacccaccatc-3 (SEQ ID NO:48)

~~-~~CaYMR212back: 5-gaatatcctttttaactcaagag-3' (SEQ ID NO: 49)

(b) 3'-and 5'-extension of this internal fragment from CaMR212

For this purpose, genomic Candida DNA libraries from Dominique Sanglard (received from RMV) were used. The YEp24 backbone of this library was used to amplify the 3'-and 5'-coding and non coding sequences with PCR. This was done by using oligos specific for the CaMR212 490 bp-fragment

(directional towards the vector flanking regions) and vector-specific oligos (directional towards the insert).

Oligos:

YEP24for (vector specific):

5'-cggaattcctatcgactacgcgatcatgg (SEQ ID NO: 35) YEP24ba (vector specific):

5'-gcgaattccgatataggcgccagcaac (SEQ ID NO: 36)

Primer YEp24for and CaMR212for gave a 500 bp fragment, encoding 5'-UTR and the 5'coding region from CaMR212.

Using primer YEp24 for and CaMR212back a 1400 bp CaMR212-fragment was amplified. Using the sequence of this 1400 bp-fragment the following new primers, specific for this fragment were designed.

Oligos:

Ca212-1: 5'-gctttcccagcaggataacattg (SEQ ID NO:50)

Ca212-2: 5'-tgagttataatgcagctgttgg (SEQ ID NO:51)

Ca212-3: 5'-catctcgtgtgaacatgattgg (SEQ ID NO:52)

Primers YEP24 for and Ca212-3 gave a 1600 bp fragment, coding for the 3'-coding region and the 3'UTS region.

With the 3 PCR fragments the 2900 bp sequence (including coding and 3'and 5'-non-coding sequences) was assembled. With the following new primers the coding sequences was amplified from genomic DNA and cloned into p413GALL-vector.

Oligos for amplifying coding region:

Ca212for: 5'-agtttcttcaacttccagatccaag (SEQ ID NO: 53)

Ca212back: 5'-gtatatttgcaactgtctctctctc (SEQ ID NO: 54)

The yeast homolog YMR212c plays a role in cell wall function because the knockout can be rescued in 1M sorbitol. In addition, YMR212c unger GAL-promoter regulation shows an increased sensitivity versus Congo Red and Calcofluor White. YMR212c is an integral membrane

protein and localizes to the plasma membrane (demonstrated by microscope analysis of YMR212-GFP fusion proteins and by biochemical analysis of YMR212-GST fusion proteins).

Example 7: CaDR325

The CaDR325 sequence is given at SEQ ID 20 and 21~~13~~.

CaDR325 was cloned based on gene fragment data from the public Stanford Candida albicans sequencing database.

(a) 3 fragments that showed homology to Saccharomyces cerevisiae YDR325 were identified, the sequences of which are disclosed in SEQ ID 22~~14~~, 23~~15~~ and 24~~16~~.

Based on these data, the following oligos were designed that allowed the verification of the database sequences and the amplification of an approx. 2200 bp internal CaDR325 fragment from genomic DNA:

cgagcatctacttgttcaaccac: hybCaYDR325ba Oligo (SEQ ID NO: 55)

gaatctctggctcgctc: 325-juls Oligo (SEQ ID NO:56)

gaccgagatacacgagaat: 325-julr Oligo (SEQ ID NO:57)

ggttaaatgatcgtgatgaat: Ca325r Oligo (SEQ ID NO:58)

caacctcactgacaaatactt: Ca325s Oligo (SEQ ID NO:59)

The finally subcloned 2200 bp internal fragment was amplified by the combination hybCaYDR325ba + 325-julr oligos.

(c) 3'-and 5'-Extension of the internal fragment: The Sanglard genomic Candida DNA library (received from RMV) in the YEP24 vector backbone was used for the amplification of 3'-and 5'-coding and non-coding sequences. This was done by using the following vector- specific oligos (directional towards the insert) and CaDR325 2200 bp fragment-specific oligos (directional towards the vector flanking sequences):

cggaattcctatcgactacgcgatcatgg (SEQ ID NO:35) : YEP24for (vector specific)

gcgaattccgatataggcgccagcaac (SEQ ID NO:36) : YEP24ba (vector specific)

acgcttccaatgtattattctcg (SEQ ID NO:60) : Oligo 1-10-A back

ggatgccaatttccctga(SEQ ID NO: 61) : Oligo 1-10-B for

catccagaagatataacggct (SEQ ID NO: 62) : Oligo 1-10-C for

tgcataatctactcagcgaca (SEQ ID NO:63) : Oligo 1-10-D back

gtggttgaacaagtagatgctcg (SEQ ID NO:64) : Oligo 1-10-E for

gcgcttgaaaccactagtgaattg(SEQ ID NO:65) : Ca325Klon 2_Fo

caattcactagtggtttcaagcgc (SEQ ID NO: 66) : Ca325Klon3Ba

The finally assembled 4700 bp sequence that included also 3'-and 5'-non-coding sequences were verified by sequencing. The coding region was subcloned into the p413RSGALL-vector.

The map is disclosed in fig. 7.

Sequences numbers are identified in field 130 of the sequence listing.

Claims:

1.-A polynucleotide having the sequence as depicted in the sequence selected from the group consisting of SEQ ID No.2,SEQ ID No.~~4~~5, SEQ ID No.~~6~~8, SEQ ID NO.~~7~~10, SEQ ID No. ~~9~~13, SEQ ID No. ~~10~~15, SEQ ID No. ~~11~~17, or SEQ ID No. ~~13~~20, homologs thereof and functional fragments thereof.

2.-The polynucleotide of claim 1 which is the gene CaNL256, homologs thereof and functional fragments thereof.

3.- The polynucleotide of claim 1 which is the gene CaBR102, homologs thereof and functional fragments thereof.

4.- The polynucleotide of claim 1 which is the gene CaIR012, homologs thereof and functional fragments thereof.

5.- The polynucleotide of claim 1 which is the gene CaMR212, homologs thereof and functional fragments thereof.

6.- The polynucleotide of claim 1 which is the gene CaDR325, homologs thereof and functional fragments thereof.

7.- The polynucleotide of claim 1 which is the gene CaOR110, homologs thereof and functional fragments thereof.

8. -The polynucleotide of claim 1 which is the gene CaJL039, homologs thereof and functional fragments thereof.

9.-A protein encoded by the polynucleotide according to claim 2 or a functional polypeptide fragment thereof.

10.-A protein encoded by the polynucleotide according to claim 3 or a functional polypeptide fragment thereof.

11. -A protein encoded by the polynucleotide according to claim 4 or a functional polypeptide fragment thereof.

12. -A protein encoded by the polynucleotide according to claim 5 or a functional polypeptide fragment thereof.

13. -A protein encoded by the polynucleotide according to claim 6 or a functional polypeptide fragment thereof.

14. -A protein encoded by the polynucleotide according to claim 7 or a functional polypeptide fragment thereof.

15. -A protein encoded by the polynucleotide according to claim 8 or a functional polypeptide fragment thereof.

16.-A plasmid deposited at the CNCM with the accession number I-2065.

17.-A plasmid deposited at the CNCM with the accession number I-2063.

18.-A plasmid deposited at the DSMZ with the accession number DSM 12977.

19.-A plasmid deposited at the DSMZ with the accession number DSM 12976.

20.-A plasmid deposited at the DSMZ with the accession number DSM 12978.

21.-A plasmid deposited at the DSMZ with the accession number DSM 12979.

22.-An ~~antiboby~~ antibody directed against the protein of claim 9 or a functional polypeptide fragment thereof.

23.- An antibody~~antiboby~~ directed against the protein of claim 10 or a functional polypeptide fragment thereof.

24.- An antibody~~antiboby~~ directed against the protein of claim 11 or a functional polypeptide fragment thereof.

25.- An antibody~~antiboby~~ directed against the protein of claim 12 or a functional polypeptide fragment thereof.

26.- An antibody~~antiboby~~ directed against the protein of claim 13 or a functional polypeptide fragment thereof.

27.- An antibody~~antiboby~~ directed against the protein of claim 14 or a functional polypeptide fragment thereof.

28.-An antibody~~antiboby~~ directed against the protein of claim 15 or a functional polypeptide fragment thereof.

29.-·A polypeptide obtainable by the process comprising the following steps:

(i) Selecting an essential gene from Saccharomyces cerevisiae

(ii) Comparing the sequence of said gene with Candida Albicans genome sequences;

(iii) Deducing homologous oligonucleotides regions

(iv) PCR amplifying the thus-obtained oligonucleotides;

(v) Using the amplimers of step (iv) for detecting the complete gene of interest; and homologs thereof and functional fragments thereof.

30.- The polynucleotide of claim 29, in which step (v) is comprised of the step of using the amplimers of step (iv) as a probe for detecting the complete gene of interest from a Candida ~~Albicans~~ albicans genomic library.

31. - The polynucleotide of claim 29, in which step (v) is comprised of the step of using the amplimers of step (iv) as a probe for detecting the complete gene of interest from a Candida ~~Albicans~~ albicans cDNA library.

32. The polyonucleotide of claim 29, in which step (v) is comprised of the step of 3′ and 5′ extension of the amplimer using a PCR method.

33.- A method for the screening of antimycotic substances wherein an essential gene from mycetes or a functionally similar gene from another pathogenic mycete, or the corresponding encoded protein, is used as target and wherein the essential gene is selected from the group consisting of CaNL256, CaBR102, CaMR212, CaDR325, Ca OR110, CaJL039, homologs thereof and functional fragments thereof.

34.- The method of claim 33 wherein mycete cells which express the essential gene, or a functionally similar mycete gene, to a different level are incubated with the substance to be tested and the growth inhibiting effect of the substance is determined.

35.- The method of claim 33 wherein said target gene or the corresponding target encoded protein is contacted in vitro with the substance to be tested and the effect of the substance of the target is determined.

36. - The method according to claim 33 wherein the screened substances partially or totally inhibit the functional expression of the essential genes or the functional activity of the encoded proteins.

37. - The method according to claim 33 wherein the screened substances partially or totally inhibit the activity of dihydropneopterin aldolase (DHNA).

38. - The method according to claim 33 wherein the screened substances partially or totally inhibit the activity of dihydropteroate synthase (DHPS).

39. - The method according to claim 33 wherein the screened substances partially or totally inhibit the activity of 7, 8-dihydro-6-hydroxymethylpterin-pyrophoshokinase (HPPK).

40. - The method according to claim 33 wherein the mycete species are selected from the group comprising Basidiomycetes, Ascomycetes and Hyphomycetes.

41. -The method according to claim 33 wherein said functionally similar genes are essential genes for Candida Spp, or Aspergillus Spp.

42. -The method according to claim 41, wherein said functionally similar genes are essential genes from Candida albicans, or Aspergillus fumigatus.

43. - A kit for diagnosis of fungal infections comprising a gene selected from the group consisting of CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 and CaJL039, a functionally similar gene thereof, a functional fragment thereof, the corresponding encoded protein or a functional polypeptidic fragment thereof, or an antibody directed against the protein encoded by the gene selected from the group consisting of CaOR110, CaMR212, CaNL256, CaBR102, CaIR012, CaDR325 and CaJL039, or by a functionally similar gene, or a polypeptidic fragment thereof.

ABSTRACT

The present invention concerns essential genes from C. albicans and their use in a method for the screening of antimycotic substances.

Fig.1

```
                 .         .         .         .
  1 ............................IHPISAESLHSHLQQLINDKPQ 22   SEQIDNO:67
                                :||:.|| :  ||.||  . .
451 PDLNIPHPRMLERTFVLEPLCELISPVHLHPVTAEPIVDHLKQLYDKQHD 500  SEQIDNO:68
                 .         .         .         .         .
 23 ETV........QESSDLLQFIPVSRLPVKDNILKFDQINHKSPTLIMGIL 64
    |            |     .|:             |            ||| || |
501 EDTLWKLVPLPYRSGVEPRFLKFKTATKLDEFTGETNRITVSPTYIMAIF 550
                 .         .         .         .         .
 65 NMTPDSFSDGGKHFG...KELDNIVKQA.EKLVSEGATIIDIGGVSTRPG 110
    | |||||||||.||       :|..|:|   : |     . |||:|| ||||
551 NATPDSFSDGGEHFADIESQLNDIIKLCKDALYLHESVIIDVGGCSTRPN 600
                 .         .         .         .         .
111 SVEPTEEEELERVIPLIRAIRQS........................... 133
    |:: .||||: | ||||:|||:|
601 SIQASEEEEIRRSIPLIKAIRESTELPQDKVILSIDTYRSNVAKEAIKVG 650
```

Fig.2

```
                         .         .         .         .         .
251 NDLNEVLDQCTKIAEKRLQLQDQIDQERQGNFNNVESHSNSPALLPPLKA 300  SEQ ID NO:69
                                                   : .
  1 .............................................KSIQL 5    SEQ ID NO:70

                         .         .         .         .         .
301 GQNGNLMRRDRSSVLILEKFWDTELDQLFKNVEGAQKFINSTKGRHILMN 350
    |   |  ::||||:::|.| ||.:|  |||.|:|| ||:      |||.
  6 GIPSN.KKKDRSSIMVLKKMWDSQLQSLFKHVDGASKFVQPLPNRHIVAE 54

                         .         .         .         .         .
351 SANWMELNTTTGKPLQMVQIFILNDLVLIADK...SRDKQNDFIVSQCYP 397
    |  | |.|     ||     :|| |||:||| |   |  .:      |
 55 SGRWFEVNVGNWKPSYPTHLFIFNDLILIAVKKSSSSSQEPTTGGSNGGS 104

                         .         .         .         .         .
398 LKDVTVTQEEFSTKRLLFKFSNSNSSLYECRDADECSRLLDVI..RKAKD 445
      .  |   |. | .  .     .    . |: |  :      |
105 KSRLQAVQCWPLTQVSLQQIKSPKKDDDKMYFINLKSKSLSYVYLTDRYD 154

                         .         .         .         .         .
446 DLCDIFHVEEENSKRIRESFRYLQSTQQTPGRENNRSPNKNK..RRSMGG 493
       :    .   .:| | | |   .|   |  |  . :  | |:
155 HFVKVTEAFNKGRNEMIQSERLLDSRLSSPSNNNGDSKEEKRQLRESLRN 204

                         .         .         .         .         .
494 SITPGRNVTGAMDQYLLQNLTLSMHSRPRSRDMSSTAQRLKFLDEGVEEI 543
    |     ||         .|
205 SGNYKEGVTDDAGGAATG*VT............................. 225
```

FIG.3

```
301 ACCCATTGCTGAAATGTTGGACTTGAAGATTGCTTTAGAAAGAGGAGTTGGTGAATGGTT SEQ ID NO:72

  0 ------------------------------------------------------------

361 TCGTAAAAATAGAGATACCAAACCAGTTCCCGGTGATTACACACAATTGAGAACATTTTT SEQ ID NO:72

  0 ------------------------------------------------------------

421 CGATAAATTATTGATCGATGAAGATACTTGGCCAAGAGATAACTTAAATGTTATACCTAA SEQ ID NO:72

  0 ------------------------------------------------------------

481 TATTGAAGGAGAAGATTATGATGAAATCTACGATCGTGCCAAATTGTTTTGGAAAAAGTT SEQ ID NO:72

  0 --------------------------------TTAAATATGTGTTGATAGTTACACATGC SEQ ID NO:71
                                    ||||| ||||||||||||||||||||||
541 TATTCCTGAATTTGAAAAGAAATTCCCCGAAATTAAAAATGTGTTGATAGTTACACATGC SEQ ID NO:72

 29 AGCAACGAAAATTGCTTTAGGATCAGCTTTATTACAGTTAAAATCAGTTACTGATGTTAT SEQ ID NO:71
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601 AGCAACGAAAATTGCTTTAGGATCAGCTTTATTACAGTTAAAATCAGTTACTGATGTTAT SEQ ID NO:72

 89 AGATGATAATCAAACTGTGTTACGTGCTGGTGCATGTTCATTATCCAAATTTGTTAGAGA SEQ ID NO:71
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
661 AGATGATAATCAAACTGTGTTACGTGCTGGTGCATGTTCATTATCCAAATTTGTTAGAGA SEQ ID NO:72

149 TGGCGAAGATAAAACCAATCATACTATTCAATGGAAAATTGTCATGAATGGTAATTGTGA SEQ ID NO:71
    ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
721 TGGCGAAGATAAAACCAATGATACTATTCAATGGAAAATTGTCATGAATGGTAATTGTGA SEQ ID NO:72

209 ATTCTTGACACAGGGTGAAGAAATGAAT--------------------------------  SEQ ID NO:7
    |||||||||||||||||||||||||||
781 ATTCTTGACACAGGGTGAAGAAATGAACTGGGATTTCCGTCGTGGTGTTGAAGCCGGGTC SEQ ID NO:7
```

FIG.4

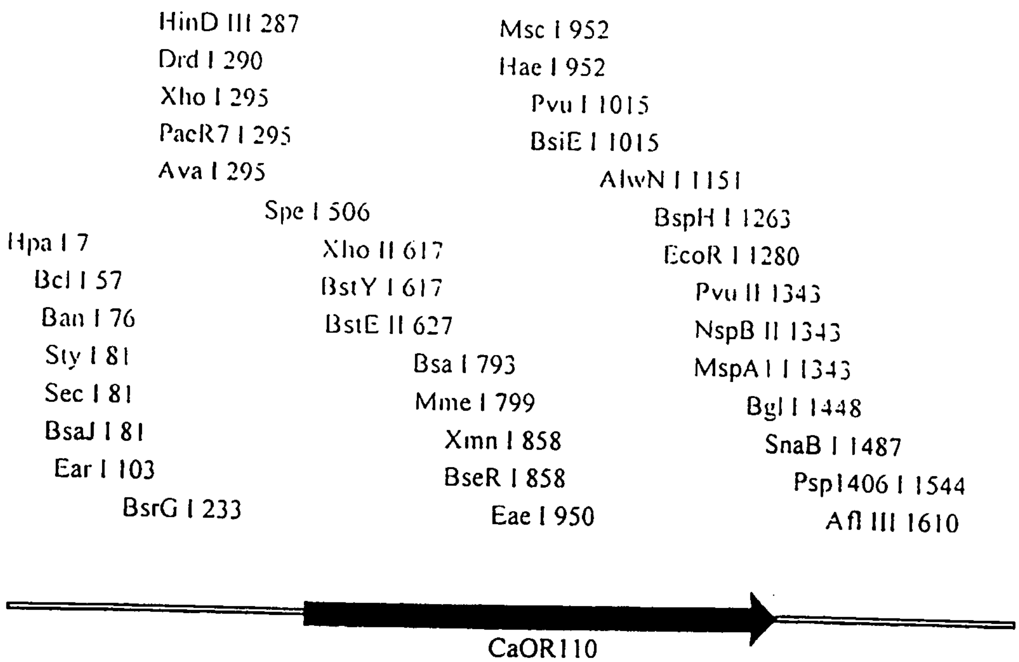

HinD III 287
Drd I 290
Xho I 295
PacR7 I 295
Ava I 295
Spe I 506
Hpa I 7
Bcl I 57
Ban I 76
Sty I 81
Sec I 81
BsaJ I 81
Ear I 103
BsrG I 233
Xho II 617
BstY I 617
BstE II 627
Bsa I 793
Mme I 799
Xmn I 858
BseR I 858
Eae I 950
Msc I 952
Hae I 952
Pvu I 1015
BsiE I 1015
AlwN I 1151
BspH I 1263
EcoR I 1280
Pvu II 1343
NspB II 1343
MspA1 I 1343
Bgl I 1448
SnaB I 1487
Psp1406 I 1544
Afl III 1610
CaOR110

FIG.5

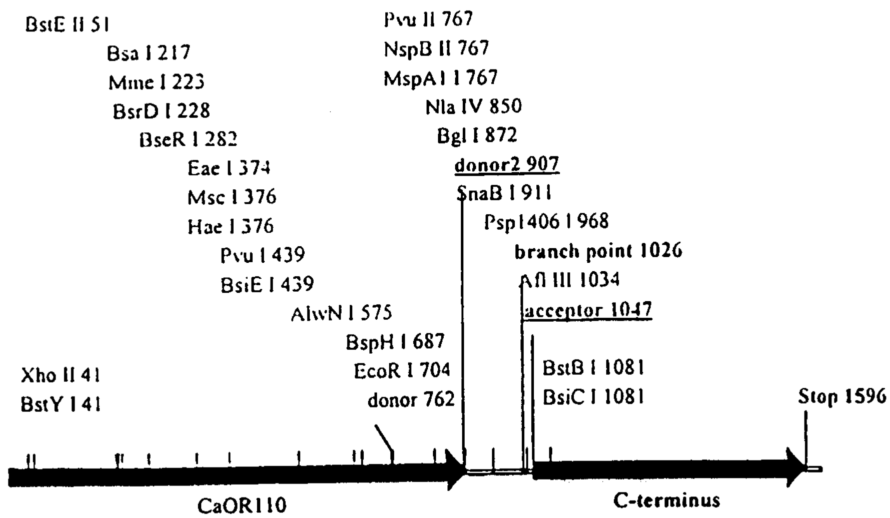

BstE II 51
Bsa I 217
Mme I 223
BsrD I 228
BseR I 282
Eae I 374
Msc I 376
Hae I 376
Pvu I 439
BsiE I 439
AlwN I 575
BspH I 687
EcoR I 704
donor 762
Xho II 41
BstY I 41
Pvu II 767
NspB II 767
MspA1 I 767
Nla IV 850
Bgl I 872
donor2 907
SnaB I 911
Psp1406 I 968
branch point 1026
Afl III 1034
acceptor 1047
BstB I 1081
BsiC I 1081
Stop 1596
CaOR110
C-terminus

FIG. 6A

6A

```
  1 ATGACGATTGAAACTATTTATATCGCAAGACACGGTTATAGATCCAATTGGTTACCACCA   60 SEQ ID NO: 13
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ATGACGATTGAAACTATTTATATCGCAAGACACGGTTATAGATCCAATTGGTTACCACCA   60 SEQ ID NO: 16

 61 CCACACCCACCAAATCCTACTGGTATTGACAGTGACCCGGCTTTAGCACCACATGGTGTT  120
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 61 CCACACCCACCAAATCCTACTGGTATTGACAGTGACCCGGCTTTAGCACCACATGGTGTT  120

121 GAACAAGCCCAACAGTTAGCTGCCTATCTTACATCATTACCTACACATGAAAAGCCTGAA  180
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 GAACAAGCCCAACAGTTAGCTGCCTATCTTACATCATTACCTACACATGAAAAGCCTGAA  180

181 TTTATTATTGCTTCACCTTTTATCGTTGTATAGAAACGTCGAGACCCATTGCCGAAATG  240
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 TTTATTATTGCTTCACCTTTTATCGTTGTATAGAAACGTCGAGACCCATTGCCGAAATG  240

241 TTGGACTTGAAGATTGCTTTAGAAAGAGGAGTTGGTGAATGGTTTCGTAAAAATAGAGAT  300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 TTGGACTTGAAGATTGCTTTAGAAAGAGGAGTTGGTGAATGGTTTCGTAAAAATAGAGAT  300

301 ACCAAACCAGTTCCCGGTGATTACACACAATTGAGAACATTTTTCGATAAATTATTGATC  360
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 ACCAAACCAGTTCCCGGTGATTACACACAATTGAGAACATTTTTCGATAAATTATTGATC  360

361 GATGAAGATACTTGGCCAAGAGATAACTTAAATGTTATACCTAATATTGAAGGAGAAGAT  420
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 GATGAAGATACTTGGCCAAGAGATAACTTAAATGTTATACCTAATATTGAAGGAGAAGAT  420

421 TATGATGAAATCTACGATCGTGCCAAATTGTTTTGGAAAAAGTTTATTCCTGAATTTGAA  480
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421 TATGATGAAATCTACGATCGTGCCAAATTGTTTTGGAAAAAGTTTATTCCTGAATTTGAA  480

481 AAGAAATTCCCCGAAATTAAAAATGTGTTGATAGTTACACATGCAGCAACGAAAATTGCT  540
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
481 AAGAAATTCCCCGAAATTAAAAATGTGTTGATAGTTACACATGCAGCAACGAAAATTGCT  540

541 TTAGGATCAGCTTTATTACAGTTAAAATCAGTTACTGATGTTATAGATGATAATCAAACT  600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
541 TTAGGATCAGCTTTATTACAGTTAAAATCAGTTACTGATGTTATAGATGATAATCAAACT  600

601 GTGTTACGTGCTGGTGCATGTTCATTATCCAAATTTGTTAGAGATGGCGAAGATAAAACC  660
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601 GTGTTACGTGCTGGTGCATGTTCATTATCCAAATTTGTTAGAGATGGCGAAGATAAAACC  660

661 AATCATACTATTCAATGGAAAATTGTCATGAATGGTAATTGTGAATTCTTGACACAGGGT  720
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

6B


```
 661 AATCATACTATTCAATGGAAAATTGTCATGAATGGTAATTGTGAATTCTTGACACAGGGT  720

 721 GAAGAAATGAACTGGGATTTCCGTCGTGGTGTTGAAGCCGGGTCAGCTGAAGATATAGCG  780
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 721 GAAGAAATGAACTGGGATTTCCGTCGTGGTGTTGAAGCCGGGTCAGCTGAAGATATAGCG  780

 781 CAAAGAAAGGCAGCAGCAGAAGCAGAAGCAAAAGCATTGAAGAAAAATGAACAAACCAAA  840
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 781 CAAAGAAAGGCAGCAGCAGAAGCAGAAGCAAAAGCATTGAAGAAAAATGAACAAACCAAA  840

 841 TCCGATGGTCCCATCACTGAATCTGCCACTGGGGCAGAAATAGATGGGAATGAAGATGAA  900
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 841 TCCGATGGTCCCATCACTGAATCTGCCACTGGGGCAGAAATAGATGGGAATGAAGATGAA  900

 901 TTTGAAGTACGTAAAACTTGAAAGAGATATTAAATAGACACAAACTTAGAAAATATAGAG  960
     ||||||
 901 TTTGAA------------------------------------------------------  906

 961 ATACAAACGTTTTGAATTTCTTGATTCACTTTTTTGTTTAAAAATAAAAATAGTTCAAAA 1020

 906 ------------------------------------------------------------  905

1021 TGAAATACTAACACATGTGTTTTTAGACATTTTATGTAACCATCGATATACCTTCAATTT 1080
                               |||||||||||||||||||||||||||||||||||
 906 --------------------------ACATTTTATGTAACCATCGATATACCTTCAATTT  939

1081 CGAATAAAATCGACAATGAAGAAGAACCACCATCAAGGACAGGTCAAGCTCCAAAATTCA 1140
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 941 CGAATAAAATCGACAATGAAGAAGAACCACCATCAAGGACAGGTCAAGCTCCAAAATTCA 1000

1141 AAAACAATATTATCAAGCCTTCAGCACAACTCCAATTTACTGATTTAAAAGAAGATCATC 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1001 AAAACAATATTATCAAGCCTTCAGCACAACTCCAATTTACTGATTTAAAAGAAGATCATC 1060

1201 CATTAGTAAAAATATCGAACAATACTATATCTGCTCAAGGCTCGTCGTCGTCGTCGTTAT 1260
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1061 CATTAGTAAAAATATCGAACAATACTATATCTGCTCAAGGCTCGTCGTCGTCGTCGTTAT 1120

1261 CAGCGTCGAAAAATGGATTTAATAGTCATACTCACAATTCAGGAGTCATTGATCCATCAG 1320
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1121 CAGCGTCGAAAAATGGATTTAATAGTCATACTCACAATTCAGGAGTCATTGATCCATCAG 1180

1321 CACTTATAGATGGGAAAATTTATCAGACTGATTGGAATCAATTACAAGGTACTGAACTAA 1380
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1181 CACTTATAGATGGGAAAATTTATCAGACTGATTGGAATCAATTACAAGGTACTGAACTAA 1240

1381 TATTTGATGAAAATGGTCAATTTATAGGCAAGGTTAAGGAACATTTGACTTGCAATAATA 1440
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1241 TATTTGATGAAAATGGTCAATTTATAGGCAAGGTTAAGGAACATTTGACTTGCAATAATA 1300
```

6C

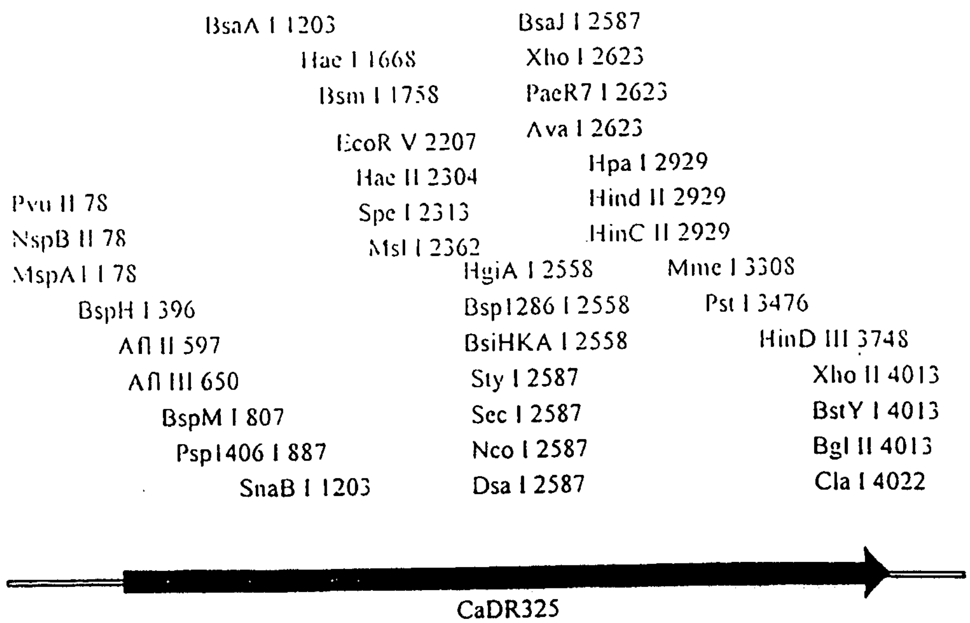

```
                      .         .         .         .         .         .
1441 ACACAAAATTCACATTAAAAAAGGCAGAAGAAGTAGAACAACTTCGTTCAGCAGATGATT 1500
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1301 ACACAAAATTCACATTAAAAAAGGCAGAAGAAGTAGAACAACTTCGTTCAGCAGATGATT 1360
                      o         o         o         o         o         o

                      .         .         .         .         .         .
1501 CTATCATGGATATAGATCAAGACTCACAAGGACAACAACCAGCTAGAAGTCAGTTCTTAA 1560
     |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
1361 CTATCATGGATATAGATCAAGACTCACAAGGACAACAACCAGCTAGAAGTCAGTTCTTAA 1420
                      o         o         o         o         o         o

                      .         .         .         .         .         .
1561 AAAGAGCAATTGTGGCTGCTAGAGCCAAAGGTAAATAAATGCTATTTTGTTATTATTATA 1620
     ||||||||||||||||||||||||||||||||||
1421 AAAGAGCAATTGTGGCTGCTAGAGCCAAAGGTAA-------------------------- 1454
                      o         o         o

   ------------------------------------------------------------
```

FIG. 7


BsaA I 1203
Hae I 1668
Bsm I 1758
EcoR V 2207
Hae II 2304
Spe I 2313
Msl I 2362
Pvu II 78
NspB II 78
MspA I I 78
BspH I 396
Afl II 597
Afl III 650
BspM I 807
Psp1406 I 887
SnaB I 1203
BsaJ I 2587
Xho I 2623
PaeR7 I 2623
Ava I 2623
Hpa I 2929
Hind II 2929
HinC II 2929
HgiA I 2558
Bsp1286 I 2558
BsiHKA I 2558
Sty I 2587
Sec I 2587
Nco I 2587
Dsa I 2587
Mme I 3308
Pst I 3476
HinD III 3748
Xho II 4013
BstY I 4013
Bgl II 4013
Cla I 4022
CaDR325